

Research Article

Evaluation of Spatial and Temporal Performance of Deep Learning Models for Travel Demand Forecasting: Application to Bike-Sharing Demand Forecasting

Jaehyung Lee  and Jinhee Kim 

Department of Urban Planning and Engineering, Yonsei University, Seoul 03722, Republic of Korea

Correspondence should be addressed to Jinhee Kim; kim.jinhee@yonsei.ac.kr

Received 6 January 2022; Revised 14 April 2022; Accepted 31 May 2022; Published 20 June 2022

Academic Editor: Rosa G. González-Ramírez

Copyright © 2022 Jaehyung Lee and Jinhee Kim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning approaches are widely employed for forecasting short-term travel demand to respond to real-time demand. Although it is critical for demand forecasting to be evenly distributed in the spatial and temporal views to support real-time mobility service operations, in related studies, the predictive performance of models has been evaluated only in terms of aggregated errors. Therefore, the present study was conducted to investigate the distribution of errors to explore spatiotemporal correlations. Six deep learning models with the same architecture, except for the base module, consisting of three stacked layers, were constructed. These models were used to forecast demands for a station-based bike-sharing service in Seoul, South Korea. To attain our goals, global and local Moran's I of the errors was introduced to evaluate the spatial and temporal performances of the deep learning approaches. The results showed that the model with convolutional long short-term memory layers, which are effective at predicting spatiotemporal data, outperformed the other models in terms of aggregated performance. However, the global Moran's I of the errors in the model reflects spatial dependency over the regions. This suggests that the best predictive performance of the model does not necessarily imply that it performs well in demand forecasting in all regions. Furthermore, cluster and outlier analyses of the errors indicated that excessive or insufficient predictions were clustered or dispersed throughout the regions. These results can be used to enhance the model by introducing the spatial correlation index into the loss function or by incorporating additional features for handling spatial correlations.

1. Introduction

Mobility services are evolving from static and prescheduled services to flexible and real-time demand-responsive services. Emerging mobility services attempt to respond to real-time demand by adjusting the route, time schedule, rate, and capacity in real time. In other words, to satisfy real-time demand, the supply of a mobility service should be dynamically replanned based on short-term demand forecasting, which requires ultra-high-resolution temporal and spatial outputs with high accuracy. In this context, deep learning approaches are widely employed for forecasting short-term travel demand.

Specifically, various deep learning approaches have been proposed to handle the features of travel demand: temporal,

spatial, and exogenous dependencies [1]. Because travel demand depends on individual activities, the daily routines of commuters shape traffic flow patterns. Therefore, temporal dependency is considered in demand prediction to cope with time-series patterns. Furthermore, spatial dependencies can be reasonably handled, given that the distance between two regions is related to the spatial similarity between regions. Moreover, exogenous features, such as weather and events, must be considered as they may have a significant impact on travel demand.

The recurrent neural network (RNN) structure provides a deep learning approach specialized in processing sequential data. RNNs can handle temporal dependencies because of their recursive structures, which allow past and present inputs to impact current outputs simultaneously.

Long short-term memory (LSTM) is a well-known type of RNN, which introduces various gate and cell states to control the effect of past information on current outputs. Xu et al. [2] used an LSTM approach to forecast free-floating bike-sharing production and attraction for various time intervals. Xu et al. [3] developed a stochastic deep learning model based on LSTM to predict the probability distribution of real-time taxi demand in all areas, instead of demand volumes. They observed that the predictive performance varied over the combinations of input features and concluded that the impact of explanatory variables, except pick-ups, was minimal. Vanichrujee et al. [4] suggested an ensemble model for taxi demand prediction and examined the performance of the model based on land use. They observed that a single model could not ensure the best performance in all areas. The convolutional neural network (CNN) considers the locality of a spatial dataset by introducing sliding convolutional filters. Zhang et al. [5] developed a real-time crowd flow forecasting system called Urban Flow, which is based on DeepST and uses convolution layers to predict various travel demands. Zhang et al. [1] enriched the DeepST model by employing a residual neural network structure [6], which is highly effective in image classification. Then, they evaluated the model by applying it to two different datasets: taxi GPS data in Beijing and the bike-sharing system in New York. Lin et al. [7] proposed a graph convolutional neural network approach to forecast station-based bike-sharing demand and compared a data-driven graph filter with the demand correlation matrix. To forecast taxi demand, Yao et al. [8] developed a DMVST-Net framework comprising an LSTM, a CNN, and structural embedding. They determined the relative increase in the error between weekends and weekdays to assess the performance on different days. Convolutional LSTM (Conv-LSTM) [9] is a deep learning approach that introduces a convolution structure into LSTM to cope with spatial and temporal correlations simultaneously. Ke et al. [10] proposed a fusion convolutional LSTM to handle spatiotemporal data and time-series data simultaneously to predict passenger demand for on-demand ride services. They also proposed spatial aggregated random forest algorithms for feature selection to reduce the training time with minimal performance loss. Guo et al. [11] developed a residual spatiotemporal architecture comprising a full CNN and extended Conv-LSTM to capture three types of dependencies simultaneously for forecasting travel demand through a simple procedure. They then applied the proposed model to taxi data in New York and travel data in Hai Kou, China, to prove the effectiveness of their model compared with various benchmark algorithms.

Meanwhile, some studies applied domain knowledge to the deep learning model for elaborating the deep learning model. Zhu et al. [12] developed a deep learning model based on the velocity thermogram method, which compares the velocity of the vehicles under the traffic accident with the historical average speed to determine the impact range of the accident. The suggested model predicts residual incident duration instead of total incident duration to implement real-time traffic incident management. Inspired by traffic flow theory, Liu et al. [13] proposed

DeepTSP to predict speed in spatiotemporal cells. They employed features based on domain knowledge to represent spatial and temporal characteristics in traffic flow. They conducted experiments in Berlin, Istanbul, and Moscow datasets to demonstrate that their model can forecast large scale traffic states effectively.

Although the aforementioned approaches contribute to improving the overall accuracy of demand forecasting, they have not been evaluated thoroughly in terms of their ability to handle temporal, spatial, and exogenous dependencies. Given that the mobility service is the objective of demand forecasting, the service must be evenly distributed from the perspective of users. In terms of operating mobility services, errors represent the difference between demand and supply, with predictive values being used to determine the supply. Considering that an imbalance between supply and demand can reduce the efficiency of the mobility service, uneven distribution of errors can lead to inefficiency under certain conditions due to excessive or insufficient supply. In other words, relatively large errors in specific regions can indicate inequity in mobility services, which is an important factor affecting the efficiency of operating mobility services. Therefore, the spatial distribution of demand should be considered when developing a demand prediction model.

Accordingly, this study was conducted to evaluate the performance of various deep learning architectures with respect to omitting temporal, spatial, and exogenous correlations in the context of travel demand forecasting. To achieve our goals, we constructed several deep-learning structures for demand forecasting for a station-based bike-sharing service in Seoul, South Korea. We employed global and local Moran's I to evaluate the spatial and temporal distribution of the errors of the deep learning architecture. These indicators are advantageous in that they provide a statistical evaluation of parameters. In addition, the variation in the global Moran's I with time and over days of the week was to explore the temporal correlations. To the best of our knowledge, the present study is the first to suggest a statistically testable index for measuring spatial dependencies in a deep learning approach. The remaining of this paper is organized as follows. Section 2 presents the preliminaries, including the variable definitions and problem formulation. Section 3 describes the deep learning model. Section 4 describes the experimental setup, and Section 5 compares the different deep learning models, focusing on spatial and temporal dependencies. Finally, Section 6 concludes the paper.

2. Preliminaries

2.1. Variable Definition. In this section, the mathematical notations and problem formulation for predicting the bike-sharing demand at each docking station are provided. D_t^l is the intensity of demand at the t th time interval at station l ; it is defined as the rent volume within 1 hour and can be expressed as follows:

$$D_t^l = \left| \left\{ \mathcal{Y}_{(i,j)} \in \mathbb{Y} : i \in T_t \wedge j \in L_l \right\} \right|, \quad (1)$$

where $|\cdot|$ represents the cardinality of the set, $y_{(i,j)}$ represents one historical data point of renting a bike at time i and location j , which belongs to all rent datasets \mathbb{Y} , and T_t and L_l denote the t th time interval and the l th station location for time and area of interest T and L , respectively. A vector for additional explanatory variables is introduced to support the demand forecasting, which comprises the following elements.

2.1.1. Time-of-Day and Day-of-Week. Time-of-day and day-of-week were extracted from the bike-sharing rental time to explain the variation in demand patterns over time. The dummy variable h_t indicates the state of time-of-day (i.e., whether sleep hours or not) and is expressed as follows:

$$h_t = \begin{cases} 1, & \text{if } t \text{ belongs to sleep hours,} \\ 0, & \text{if } t \text{ does not belongs to sleep hours,} \end{cases} \quad (2)$$

where t represents the date on which the bike was rented. In addition, a dummy variable w_t is introduced to distinguish between the travel demand patterns on weekdays and weekends:

$$w_t = \begin{cases} 1, & \text{if } t \text{ belongs to weekdays,} \\ 0, & \text{if } t \text{ belongs to weekends.} \end{cases} \quad (3)$$

2.1.2. Station Numbers and Locations. The station number variable was used to identify specific bike stations. This variable is one-hot-encoded to transform categorical variables into dummy variables. s_n , n th component of station numbers vector of S^N , is specified as follows:

$$s_n = \begin{cases} 1, & \text{if station } n \text{ is the target of analysis,} \\ 0, & \text{if station } n \text{ is not the target of analysis.} \end{cases} \quad (4)$$

Meanwhile, the station location vector S^L , which comprises the x and y coordinates of the bike stations, is introduced to handle spatial correlations.

2.1.3. Weather. Five weather features were considered: weather state, temperature, humidity, wind speed, and precipitation. The weather state as_t vector contains four elements, which are encoded as follows:

$$as_t^l = \begin{cases} (0, 0, 0, 1), & \text{if weather is snow at time } t, \text{ station location } l, \\ (0, 0, 1, 0), & \text{if weather is sleet at time } t, \text{ station location } l, \\ (0, 1, 0, 0), & \text{if weather is rain at time } t, \text{ station location } l, \\ (1, 0, 0, 0), & \text{otherwise at time interval } t, \text{ station location } l. \end{cases} \quad (5)$$

The observations regarding temperature at_t^l , humidity ah_t^l , wind speed aw_t^l , precipitation ap_t^l at time t , and location l were used because they are continuous variables. Before the experiments, the continuous variables were processed via min-max normalization, which converts variables into values between 0 and 1.

2.2. Problem Formulation. Based on historical bike-sharing and weather data, the station-based bike-sharing demand prediction problem can be formulated as follows:

$$D_t^l = \mathcal{F}(D_{t-i}^l | i = i_1: i_d; h_{t-i}, w_{t-i} | i = i_1: i_s; as_{t-i}^l, at_{t-i}^l, ah_{t-i}^l, aw_{t-i}^l, ap_{t-i}^l | i = i_1: i_a; S^N, S^L), \quad (6)$$

where $\mathcal{F}(\cdot)$ denotes the prediction function. The subscript of interval i indicates the length of the sequences characterizing bike-sharing demand d , time slot s , and atmosphere a .

3. Methodology

Inspired by previous studies [10, 14], we developed a deep learning architecture to predict the bike-sharing demands for the stations within 1 hour. The proposed deep learning architecture is illustrated in Figure 1. Six different models

were developed with the same structures, except for the base modules; the modules comprised three identical layers with 32 units. If needed, flattened layers were introduced to transform the outputs of the modules into one-dimensional data. Two outputs from the modules were concatenated, and the concatenated output was linearly transformed to fit its dimension into the prediction values. A brief description of the layers in the modules of the six models is provided as follows.

- (1) Artificial neural network (ANN): this module consists of three fully connected hidden layers, each comprising 32 hidden units activated by a sigmoid function. The ANN model employed all the features to predict the bike-sharing demands. Past demands at the target station were used to handle temporal correlations. The past bike-sharing demands include eight historical demands intensity from one to four hours ago, one day, one week, one month, and one year ago. For other explanatory features, datasets were used at the time of the analysis.
- (2) Vanilla recurrent neural network (RNN): this module comprised three stacked vanilla RNN layers, which can handle sequential data. The bike demand and weather data were processed sequentially by the RNN model. The timestamp composition of the bike demand and weather was identical to that in the

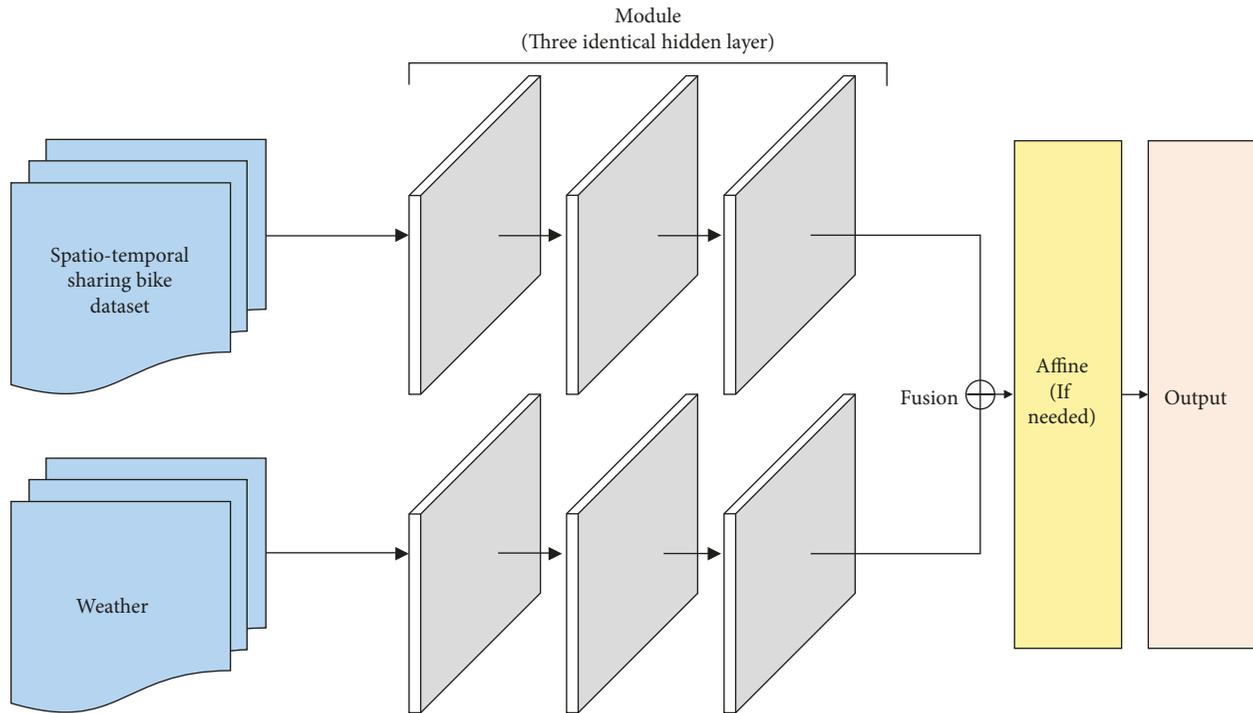


FIGURE 1: Proposed framework for prediction of bike-sharing demands.

ANN model. Meanwhile, the time steps of the time-of-day and day-of-time variables were the same, except that the time of analysis was added because they are features extracted from the data of interest.

- (3) Long short-term memory (LSTM): LSTM [15] is an RNN architecture that employs various gates and cell states to prevent the vanishing or exploding gradient problem. Because LSTM is a family of RNNs, all conditions were the same as in the RNN model, except that the layers comprising the module were LSTM.
- (4) Convolutional neural network (CNN): CNN [16] has been widely used to process spatial data. To employ a convolutional structure in the deep learning architecture, the area of interest was uniformly divided into $200\text{ m} \times 200\text{ m}$ grids. The datasets were transformed into feature maps. Because the grid already contained the station locations, station number and location features were not used in the CNN model. The length of the sequence was set to be equal to that in the ANN model.
- (5) Convolutional LSTM (Conv-LSTM): Conv-LSTM [9] was adopted to address the spatiotemporal sequence prediction problem. To overcome this problem, a convolutional structure, which is a well-known deep learning architecture for handling temporal correlations, was introduced into the LSTM structure. All inputs, hidden states, and gates in the LSTM were transformed into three-dimensional tensors. The structure of the input data was identical to that in the CNN, and the composition of the time step for features was the same as that in the LSTM.

- (6) Graph convolutional neural network (GCN): a GCN was developed to capture the locality of non-Euclidean data after noticing the advantage of CNN, which captures local features effectively. In this study, a graph convolution layer [17] was introduced among various graph-based deep learning models. Station number and location features were not used in the GCN model because a graph contains information about the relation between stations. The structure of the input data and the length of the sequence were identical to those in the CNN model. Unlike other models, an adjacent matrix was employed to represent the relation between two different stations. The adjacent matrix was identical to a spatial weight matrix, whose diagonal element is zero and nondiagonal element is the inverse of travel time when riding a bike between two stations.

4. Experiments

4.1. Site and Dataset. In the experiment, various deep learning methods were used to forecast the demand for public bike-sharing service at docking stations in Seoul. Since the launch of the service in September 2015, the cumulative utilization of public bikes has exceeded 60 million. The application site is the Seodaemun district, where the station density is intermediate among those in the 25 districts of Seoul. The population and area of the Seodaemun district are approximately 307,000 and 17.63 km^2 , respectively. The Seodaemun district has 74 bike-sharing stations, most of which are located along arterial roads and subway lines. Figure 2 depicts the application site and the locations of the bike-sharing stations.

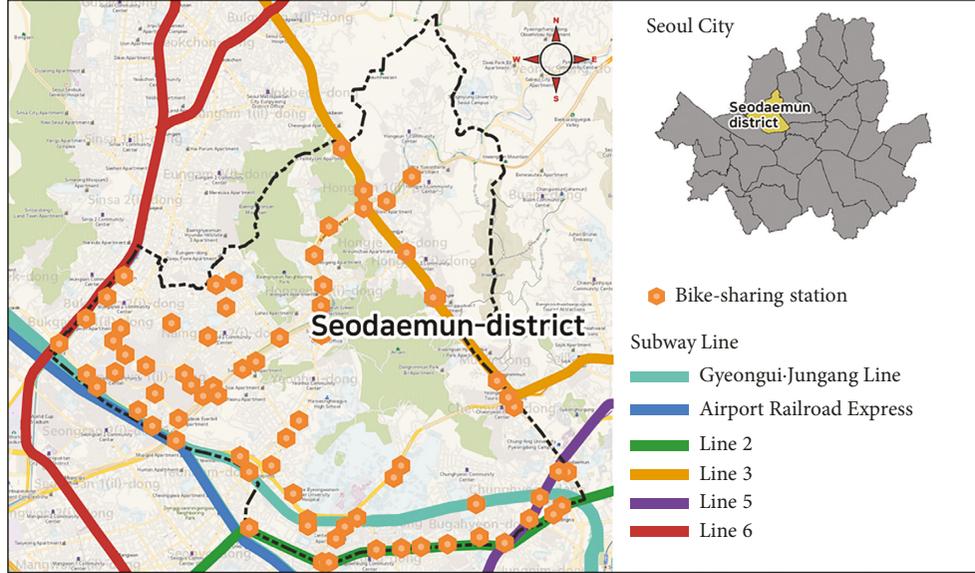


FIGURE 2: Location of bike-sharing station in Seodaemun district, Seoul, Korea.

Two different datasets were used to forecast public bike demand: the rental and return histories of bike-sharing and meteorological data, which were collected from the Seoul Open Data Plaza and National Climate Data Center, respectively. The period of the two datasets was from January 1, 2017, to December 31, 2020. Table 1 represents the detailed description of sharing bike demand dataset. Meanwhile, some missing values in the weather dataset were imputed using K-nearest neighbor algorithms with the heterogeneous Euclidean overlap metric distance [18] to handle both discrete and continuous variables simultaneously.

4.2. Settings. To ensure fairness, all the experiments were conducted under the same conditions; the hardware included a workstation with an Intel(R) i9-10850K, 128 GB RAM, and NVIDIA GeForce RTX 3090. All models were developed using Keras [19], TensorFlow [20], and Spectral [21] and were trained using the Adam optimizer [22] with an early stopping mechanism. The loss function was the mean squared error (MSE), and the number of training epochs, batch size, and learning rate were set to 100, 30, and 0.001, respectively. Filter size and strides were set to 3×3 and 1 in the CNN and the Conv-LSTM models, respectively.

To test the prediction performance of the model, approximately 80% of the data from January 1, 2017, to March 22, 2020 (1,177 days) were used as the training set, and the data from March 23, 2020, to December 31, 2020 (284 days) were used as the test set. For early stopping, the training sets were divided into 75% training sets and 25% validation sets. A seed was set to split the dataset into training and validation sets that were identically independent of the model.

4.3. Performance Metrics. The mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were introduced to evaluate the suggested models:

TABLE 1: The description of sharing bike demand dataset.

Location	Seodaemun district, Seoul, Korea	
Time span	January 1, 2017–December 31, 2020	
Time interval	1 hour	
Number of stations	74	
Total bike demand	1,685,715	
Hourly bike demand (by station)	Mean	0.650
	Standard deviation	1.422
	Max	54
	Min	0

$$MAE = \frac{1}{D} \sum_l \sum_t |e_t^l|,$$

$$RMSE = \sqrt{\frac{1}{D} \sum_l \sum_t e_t^{l2}}, \quad (7)$$

$$MAPE = \frac{1}{D} \sum_l \sum_t \left| \frac{e_t^l}{D_t^l} \right|,$$

where

$$e_t^l = D_t^l - \widehat{D}_t^l, \quad (8)$$

where D_t^l and \widehat{D}_t^l denote the real and estimated values of the bike-sharing demand at station l for the t th time step, respectively, e_t^l denotes the error in region i for the t th time interval, which can be calculated as the difference between the real value D_t^l and the predicted value \widehat{D}_t^l , and D represents the number of samples. As mentioned, however, MAE cannot consider temporal and spatial correlations because it evaluates the total error without respect to the spatial and temporal distributions of the errors. To

investigate the extent to which the temporal and spatial dependencies are correctly handled, we employed global Moran's I [23, 24], which measures spatial autocorrelation. The global Moran's I is expressed as follows:

$$I_t = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (e_t^i - \bar{e}_t)(e_t^j - \bar{e}_t)}{\sum_i (e_t^i - \bar{e}_t)^2}, \quad (9)$$

where N denotes the number of spatial units, \bar{e}_t represents the mean error at time t throughout the region, and w_{ij} represents a spatial weight matrix with a diagonal fraction of zero, which can be calculated using the inverse of the travel time of a bike between regions i and j , as follows:

$$w_{ij} = \begin{cases} \frac{1}{t_{ij}}, & (i \neq j), \\ 0, & (i = j), \end{cases} \quad (10)$$

where t_{ij} is the travel time of the bike between regions i and j . In the absence of spatial correlations, the expected values $E(I^G)$ and variance $Var(I^G)$ of the global Moran's I can be expressed as follows:

$$E(I^G) = \frac{-1}{N-1}, \quad (11)$$

$$Var(I_t^G) = \frac{A-B}{(N-1)(N-2)(N-3)W^2} - (E(I^G))^2,$$

where

$$A = N[(N^2 - 3N + 3)S_1 - NS_2 + 3W^2],$$

$$B = S_3^t[(N^2 - N)S_1 - 2NS_2 + 6W^2],$$

$$W = \sum_i \sum_j w_{ij},$$

$$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2, \quad (12)$$

$$S_2 = \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2,$$

$$S_3^t = \frac{N \sum_i (e_t^i - \bar{e}_t)^4}{\left(\sum_i (e_t^i - \bar{e}_t)^2 \right)^2}.$$

The Z-score can be obtained via statistical hypothesis testing under the null hypothesis, where no spatial dependencies exist as follows:

$$Z = \frac{I - E(I^G)}{\sqrt{Var(I_t^G)}}. \quad (13)$$

Statistically significant Z-scores indicate that the regions are spatially dispersed or clustered, depending on the signs

of the statistical values. Meanwhile, the expected values $E(I^L)$ and variance $Var(I^L)$ in the absence of spatial correlations can be calculated as follows:

$$E(I^L) = \frac{-\sum_{j \neq i} w_{ij}}{N-1}, \quad (14)$$

$$Var(I_t^L) = A - B - (E(I^L))^2,$$

where

$$A = \frac{(N - b_{2_i}) \sum_{j \neq i} w_{ij}}{N-1},$$

$$B = \frac{(2b_{2_i} - N) \sum_{k \neq i} \sum_{h \neq i} w_{ik} w_{ih}}{(N-1)(N-2)}, \quad (15)$$

$$b_{2_i}^t = \frac{N \sum_i (e_t^i - \bar{e}_t)^4}{\left(\sum_i (e_t^i - \bar{e}_t)^2 \right)^2}.$$

Similar to the global Moran's I, the Z-score at region i for time step t indicates local clusters or outliers and can be measured as follows:

$$Z_i = \frac{I_t^i - E(I^L)}{\sqrt{Var(I_t^L)}}. \quad (16)$$

Meanwhile, local Moran's I [25], which is a common metric for calculating spatial association at a specific region i , is expressed as follows:

$$I_t^i = \frac{(e_t^i - \bar{e}_t)}{\sum_i (e_t^i - \bar{e}_t)^2} \sum_{j \neq i} w_{ij} \frac{(e_t^j - \bar{e}_t)}{\sum_j (e_t^j - \bar{e}_t)^2}. \quad (17)$$

5. Results

5.1. Model Comparisons. The predictive performance and spatial correlation of the six deep learning models are presented in Table 2. The results show that the model with Conv-LSTM (Model 5) outperforms the other models in terms of MAE, while the model with RNN (Model 2) outperforms the other models in terms of RMSE and MAPE. However, MAPE has a critical weakness in which MAPE should be estimated except for most datasets and overestimated when actual values are small because MAPE is calculated relative error using the sharing bike dataset with no bike demand intensity. MAE and RMSE are comparable with the exception that RMSE is more sensitive to outliers. Therefore, we focus on MAE and RMSE rather than MAPE. Meanwhile, the global Moran's I of the models employing RNN, CNN, and Conv-LSTM is positive and statistically significant. These results indicate that the mean errors of the demand forecast by these models are not randomly distributed over the study area. In other words, the errors are likely agglomerated throughout the regions.

TABLE 2: Performance comparison of the models.

Model	Hidden layer	MAE	RMSE	MAPE ^{a)}	Global Moran's I of errors (%)		
					Values	z-statistics	p value
Model 1	ANN	0.027483	0.038449	77.84	0.0042	0.9671	0.3335
Model 2	RNN	0.016859	0.026148	49.64	0.0314**	2.5700	0.0102
Model 3	LSTM	0.016213	0.026155	56.34	0.0091	1.3428	0.1793
Model 4	CNN	0.019871	0.029821	65.22	0.0642***	4.1771	0.0000
Model 5	Conv-LSTM	0.016008	0.029356	64.28	0.0645***	4.5588	0.0000
Model 6	GCN	0.017668	0.030879	65.70	-0.0023	0.6177	0.5368

Note: ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively; a), MAPE is calculated using datasets with at least one bike demand intensity.

5.2. Spatiotemporal Dependencies. To explore spatiotemporal dependencies, the average MAE, RMSE, and global Moran's I of the mean errors were evaluated by varying the time-of-day and day of week. Figure 3 shows the average MAE, RMSE, and global Moran's I for the time of day. The number of observations in each cell was 284, which is the time span of the test sets.

First, the patterns of MAE and RMSE are comparable, while they are different from those of global Moran's I. This suggests that predictive performance is not necessarily correlated with the spatial and temporal distributions of errors. For example, the MAE between 9 a.m. and 12 p.m. for the CNN model is 0.019, while the global Moran's I under the same conditions varies from 0.002 to 0.015. This indicates that MAE and global Moran's I may not be correlated. With respect to MAE and RMSE, the temporal tendency of models is similar in which the mean error is relatively high in peak hours. Meanwhile, the Conv-LSTM model, which is the best performing model in terms of overall MAE and RMSE, does not always outperform the other models, regardless of the time-of-day. The results suggest that temporal evaluation is essential because the best-performing model in terms of overall predictive performance cannot ensure the best performance at all time intervals. In terms of global Moran's I, the positive spatial correlations are notable in the CNN and Conv-LSTM models, including in the convolutional structure. The results suggest that the errors are clustered over certain regions in the CNN and Conv-LSTM models for some time intervals. Although the Conv-LSTM model outperforms all other models from 7 p.m. to 11 p.m. in terms of MAE, the RNN or LSTM model can be an alternative for demand forecasting during periods for which the positive spatial correlations of the errors are strong in the Conv-LSTM model. The results also indicate that temporal correlations should be investigated in the CNN and Conv-LSTM models, given that the time period of significant spatial dependencies for errors is fairly continuous, particularly during the early part of the night.

Figure 4 shows the average MAE, RMSE, and global Moran's I for the day-of-week. The number of observations to obtain the values for the day-of-week is 984 or 960 (i.e., every hour during 41 or 40 days), depending on the number of days of the week. The MAE and RMSE are almost stable regardless of the day of the week, although differences exist between the models. Meanwhile, distinct patterns cannot be found in the global Moran's I of the errors. In accordance

with the analysis of the time-of-day, the Conv-LSTM model does not always outperform the other models in terms of MAE and RMSE, regardless of the day of the week. For the CNN model, the MAE on weekends was better than that on weekdays, while positive spatial correlations between the errors were more evident on weekends. This suggests that the errors of the CNN model are more agglomerated on weekends, although the difference in MAE between weekdays and weekends was not significant. In other words, the CNN model achieves a slightly better MAE on weekends than on weekdays, instead of reinforcing positive spatial correlations. Given that spatially clustered errors represent prediction failure for agglomerated regions, the CNN model is unsuitable for forecasting demand on weekends. Meanwhile, no significant difference between weekdays and weekends was observed in the other models.

Figure 5 presents autocorrelation and partial autocorrelation plots of global Moran's I. Autocorrelation and partial autocorrelation can provide information about the degree of correlations between variable and delay itself. In case of the ANN model, autocorrelation is stable regardless of time lag, while autocorrelation decreases when time lag increases in other models. Meanwhile, partial autocorrelations tend to decrease as time lag increases in all models. Similar to autocorrelations, partial correlations in the ANN model are relatively higher than other models. The results imply that deep learning structure considering spatial and temporal correlations can alleviate spatio-temporal correlations compared to the ANN model.

The aforementioned results can be summarized as follows. First, it is hard to achieve minimizing errors and handling spatiotemporal correlation simultaneously, considering the pattern of performance metrics and global Moran's I are not similar. It may be because the deep learning layer aims at different goals. For example, RNN was developed to control sequential information while the convolution layer has a strength in handling spatial relations. Although the performance of the model is most important, it is possible that other objectives should be considered in the application, such as operating transport services.

Second, the convolution layer, which is well-known for processing spatial grid data, is not suitable for handling spatial correlations in this experiment. It seems that travelers cannot pass through the middle of the site by bike because it is mountainous. In other words, the travel distance is much farther than the Euclidean distance between two stations due

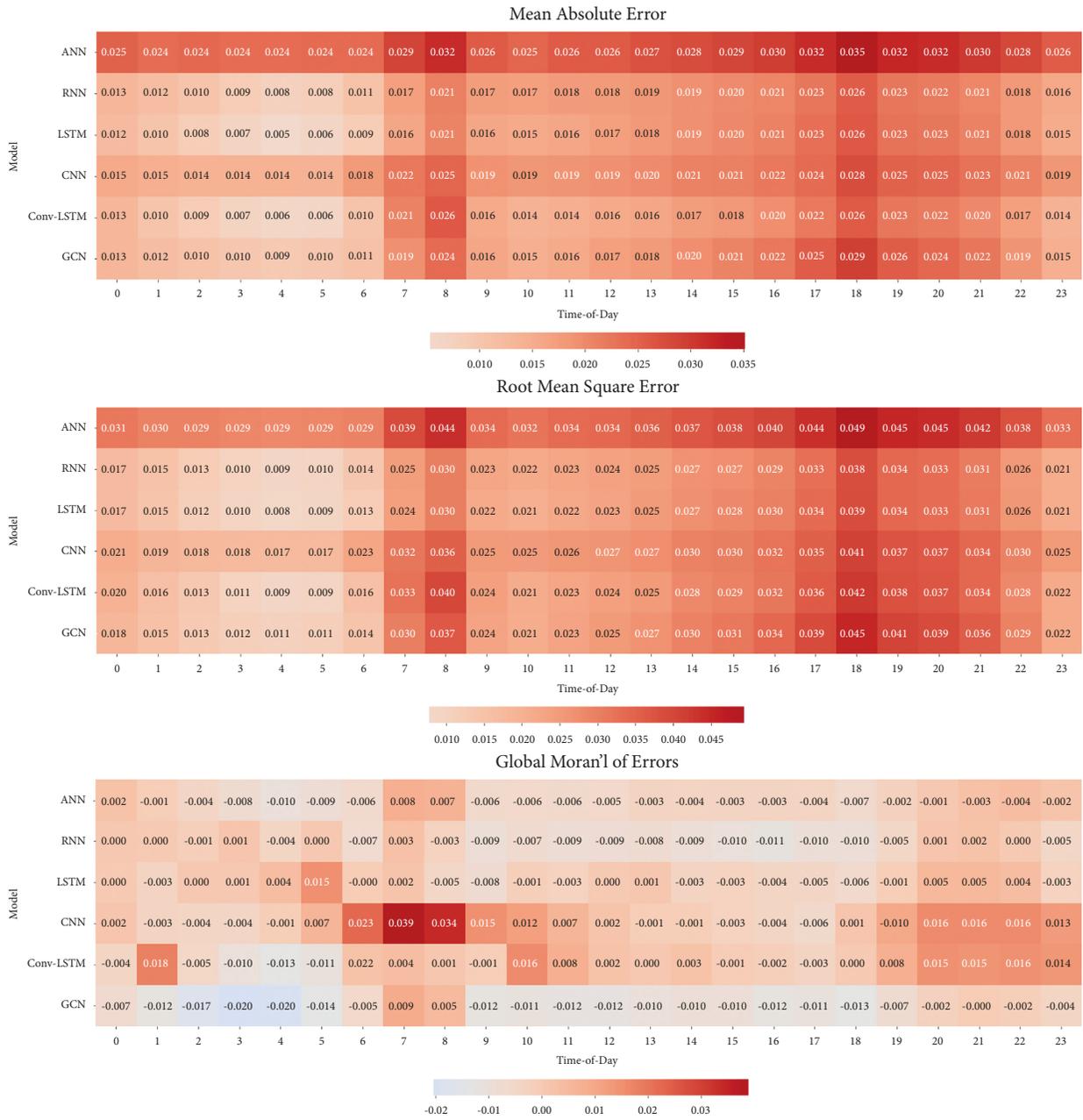


FIGURE 3: Performance metrics and global Moran's I of errors for time-of-day.

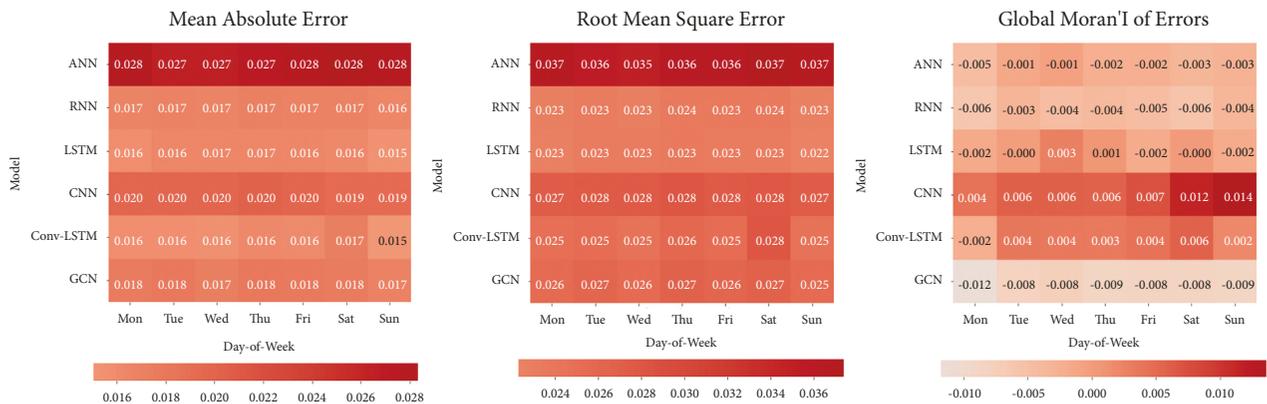


FIGURE 4: Performance metrics and global Moran's I of error for day-of-week.

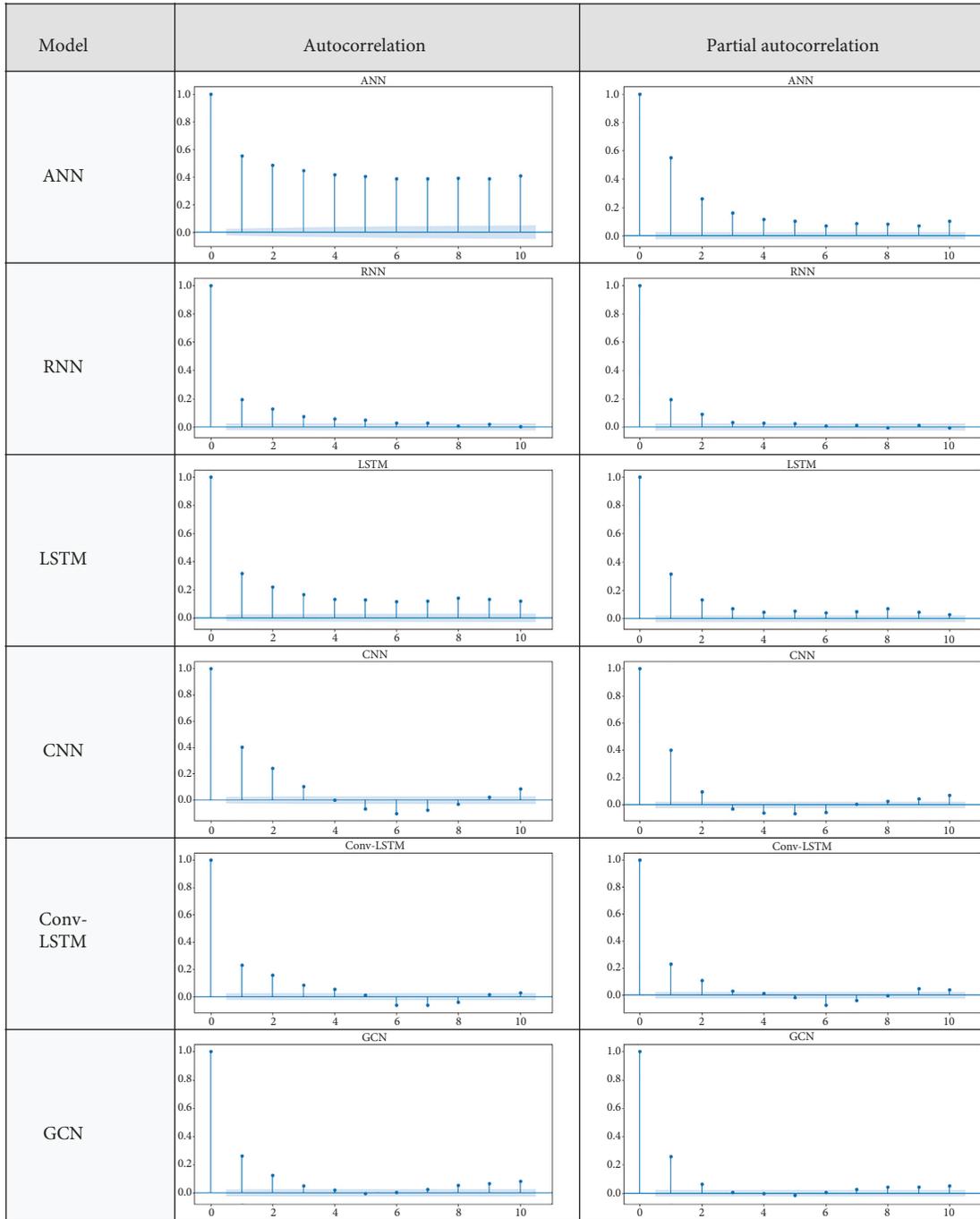


FIGURE 5: Autocorrelation and partial autocorrelation plot of global Moran’s I.

to geographical obstacles. Since spatial adjacency may be biased due to the difference between Euclidean and travel distance, striding common convolutional filters cannot capture local patterns around mountainous areas. On the contrary, there is no spatial correlation in the GCN model, which also has a convolutional structure. It seems that an adjacent matrix given as prior information can contribute to handling spatial correlation effectively.

Figure 6 shows the results of the cluster and outlier analyses for each station. Voronoi diagrams—which represent partitioning of a plane divided into a set of points

close to a specific point—were generated to separate the spatial zone based on the locations of the bike stations. The local Moran’s I was calculated using the mean error of the bike-sharing demand. Cluster and outlier analyses were conducted using bike-sharing demand data and spatial lag, which represent the product sum of spatial weight and bike-sharing demand. The results of the cluster and outlier analyses indicate two states: one is in which the values of bike-sharing demand and spatial lag at a specific station are higher than the mean value and the other is in which the aforementioned values are lower than the mean

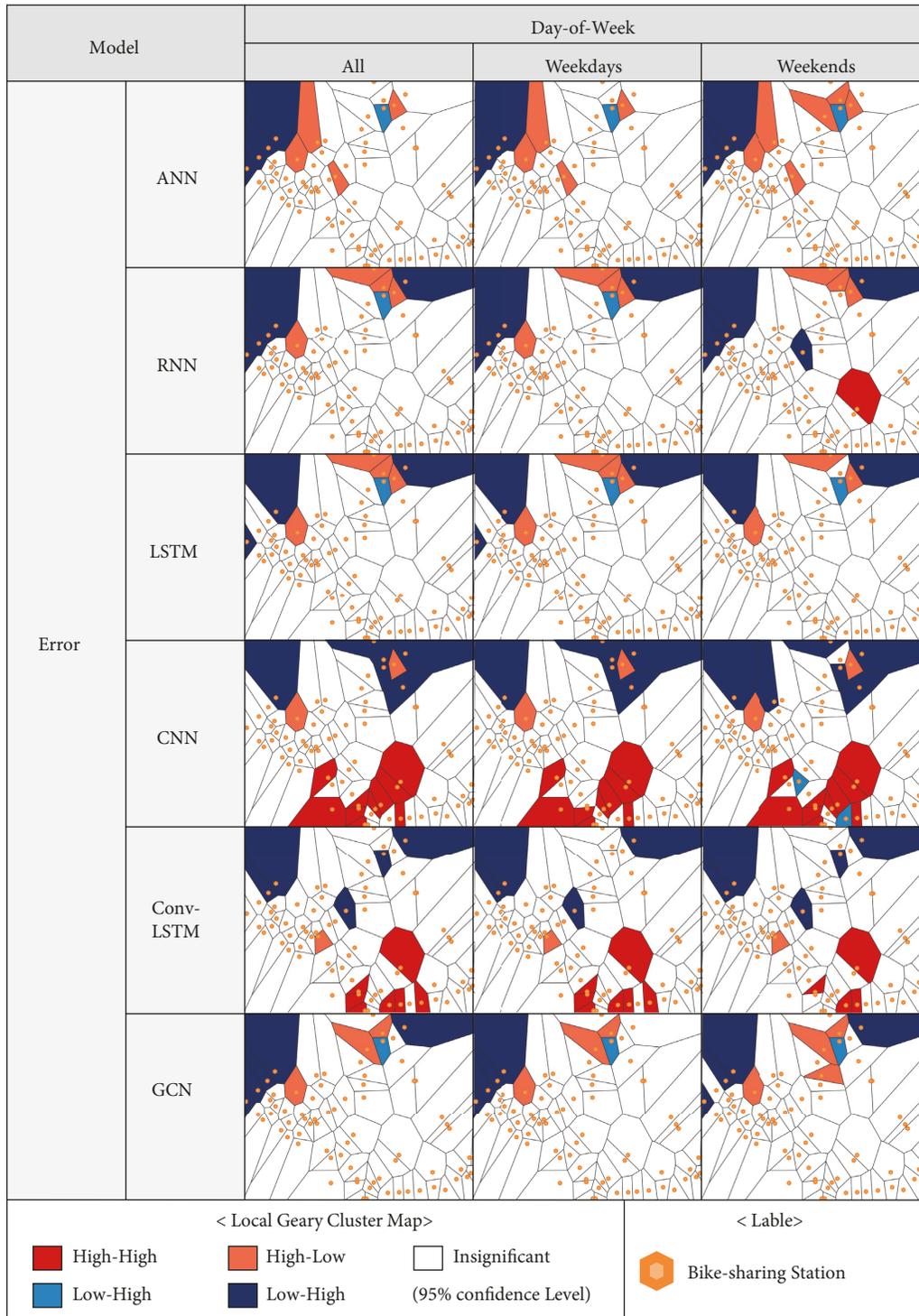


FIGURE 6: Results of cluster and outlier analyses for stations.

value. The significance represents the confidence level of the local Moran’s I, which indicates whether the error for the region is spatially clustered or dispersed. The cluster and outlier analysis provides more detailed information compared to global Moran’s I in that detecting where demand forecasting is spatially correlated.

Regardless of the model and the time of interest, a cold spot, defined as an error in bike-sharing demand that is

lower than the mean error at a specific station of interest and its surroundings, can be found in the upper left corner. By contrast, a hot spot indicates that the error is higher than the mean error at the target stations and their surroundings, and significant hotspots are observed at the bottom center stations in the CNN and Conv-LSTM models. These results suggest that the spatial correlations in the CNN and Conv-LSTM models are more prominent than those in the other

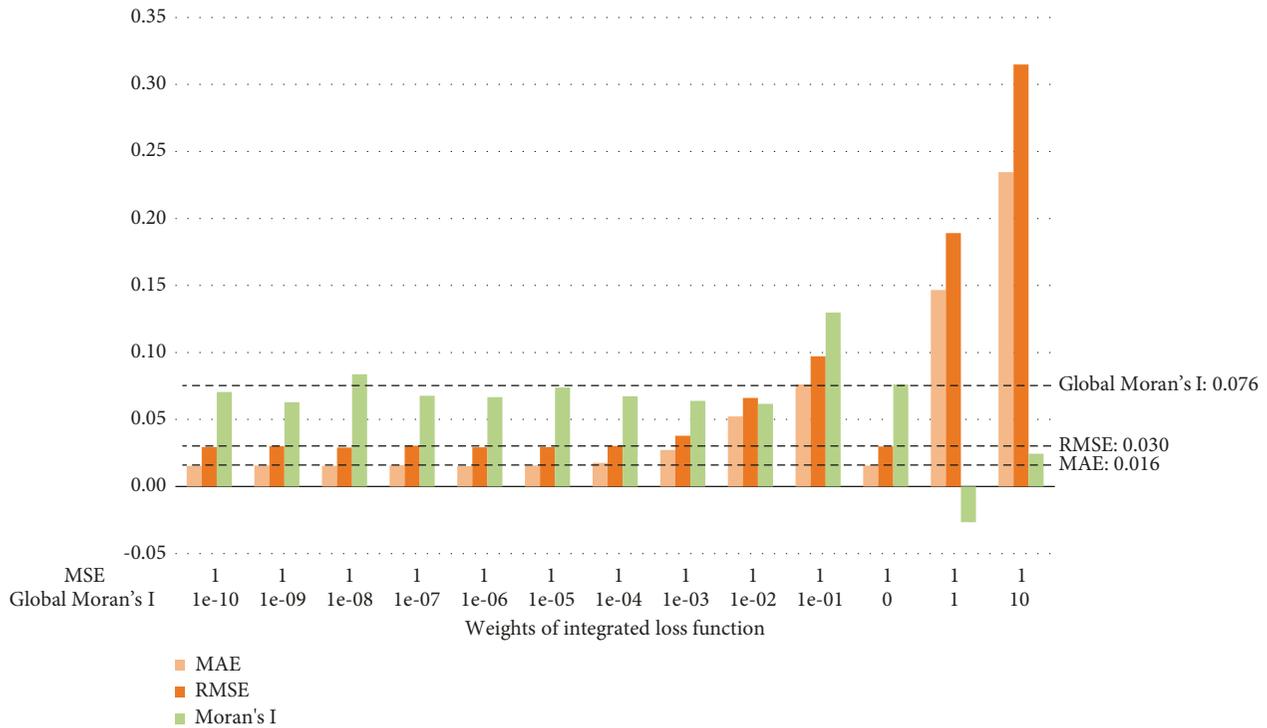


FIGURE 7: Performance metrics and global Moran's I of the Conv-LSTM model varying over-weighted value.

models; this is consistent with the variation in global Moran's I over time-of-day and day-of-week. The area where the hotspot appears in the CNN and Conv-LSTM models is the boundary of the area where bikes cannot pass through. Since the area around the hotspot has no bike-sharing demand, it may be hard to activate a convolutional filter in this area. As a result, the bike demand at bottom stations is underestimated. Meanwhile, the difference between the results of cluster and outlier analyses for weekends and weekdays is minimal in all models. This suggests that the prediction of bike-sharing demand for weekdays and weekends is similar.

5.3. Model Selection. In this experiment, two different objectives were considered: minimization sum of errors and even distribution of errors. Most of all, the accuracy of demand forecasting is critical. However, given that the demand forecasting model could be used in transportation planning, the spatial distribution of errors should also be considered in the application view. For example, if traffic demand is inaccurately predicted in a certain area, it is difficult to develop effective transportation policies or improve operation service since the premise is incorrect. When there is no significant difference in performance between the optimal model and other models, the distribution of errors could be crucial in deciding which model to use. In other words, the selection of the model is determined by which goal is more focused on.

The Conv-LSTM model is optimal in terms of accuracy, which is the most essential factor in model adoption. On the contrary, the Conv-LSTM model has a weakness in which it

contains spatial correlations. Therefore, the LSTM model could be an alternative because its performance is not significantly different from the Conv-LSTM model and there are no spatial correlations in the LSTM model.

In the same vein, establishing a loss function that integrates two different goals could be another solution for the multiobjective problem. To investigate the effect of the integrated loss function, the deep learning model with a loss function that weighted sum of mean square error and global Moran's I was trained. We choose the Conv-LSTM model because global Moran's I is significant in the Conv-LSTM model although it is the best model in terms of performance metrics. Figure 7 describes performance metrics and global Moran's I of the Conv-LSTM model varying over weighted value.

When the weighted value of global Moran's I between 0.01 to 1e-10, there is no significant difference in performance metrics compared to the model using the MSE loss function. Meanwhile, global Moran's I of the Conv-LSTM model with weighted loss function tends to be slightly lower than the Conv-LSTM model with MSE loss function. Although the integrated loss function cannot improve performance metrics significantly and exclude spatial correlations of error completely, it is significant that it can contribute to alleviating spatial correlations.

Furthermore, although the Conv-LSTM model is the best in terms of aggregated performance, it is not always optimal over time. In particular, the performance of the Conv-LSTM model is relatively low in the morning peak hour. It may be because shared parameters over time and region are biased in specific situations. Employing various demand forecasting models over time could be a solution to

this problem. Similarly, ensemble learning could provide better results because the flexibility of the model is improved.

6. Discussion and Conclusion

The present study explored the spatiotemporal dependencies of travel demands using a deep learning approach. To measure the spatial correlation, the global and local Moran's I of the errors between the predicted and real values were adopted. While the results showed that the Conv-LSTM model outperforms other models, spatial dependencies of the errors were found in the Conv-LSTM model. This indicates that the best predictive performance does not necessarily correspond to the prediction performance being evenly distributed throughout the regions. In other words, poor performance may be found in spatially clustered regions; nonetheless, the overall predictive performance is high. In addition, the result of cluster and outlier analysis reveals that the bike demand forecasting errors at bottom center stations are spatially agglomerated.

The contribution of this study lies in employing the global and local Moran's I of errors for investigating spatiotemporal correlations, which can be verified through statistical tests. Most previous studies [1, 14] have not evaluated spatiotemporal correlations, although travel demand data have both spatial and temporal correlations. Considering that errors in demand forecasting leads to an imbalance between supply and demand in mobility service operations, the distribution of errors should be investigated though the aggregated performances provide condensed power of explanation. The global Moran's I of errors can indicate whether the degree of overfitting or underfitting is clustered or dispersed, by measuring the spatial correlations of errors throughout regions. Meanwhile, local Moran's I can detect agglomerated regions with poor demand forecasting.

The global and local Moran's I suggested in this study can be applied throughout the analysis and feedback process. First, they can be used to evaluate the model after experiments in the spatiotemporal view. Global Moran's I show the degree of overall error distribution and the results cluster and outlier analysis show where demand is overestimated or underestimated. These results could contribute to elaborating models. For example, when the spatial correlation of predictive performance is identified in the model, it can be elaborated by changing the structure of the model or using additional features indicating clustered or dispersed states of errors from a different perspective. In the same vein, global Moran's I can be considered as loss function. For example, the weighted sum of predictive performance and global Moran's I is a plausible loss function that considers both performance and spatial correlations. Meanwhile, providing prior information such as an adjacent matrix may be a solution for controlling spatial correlation considering the global and local Moran's I of the GCN model were low compared to other models.

Although this study provides insight into the analysis of spatiotemporal dependencies, several factors should be

considered. First, although global Moran's I and autocorrelation were introduced to investigate spatiotemporal correlations in this study, the simultaneous measurement of spatiotemporal correlations was not verified in this study. Second, a novel model to handle spatiotemporal correlations is not provided, although this study suggests a statistically testable index to evaluate spatiotemporal performance. Further studies should consider the aforementioned issues to handle spatiotemporal dependencies for better performance and homogenous demand forecasting in the spatial and temporal domains.

Data Availability

The data can be obtained from the corresponding author upon request for academic purposes.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) and grant was funded by the Ministry of Land, Infrastructure and Transport (Grant No. 21AMDP-C161756-01).

References

- [1] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017*, San Francisco, CA, USA, pp. 1655–1661, 2019.
- [2] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 47–60, 2018.
- [3] J. Xu, R. Rahmatizadeh, L. Boloni, and D. Turgut, "Real-Time prediction of taxi demand using recurrent neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, 2018.
- [4] U. Vanichrujee, T. Horanont, W. Pattara-atikom, T. Theeramunkong, and T. Shinozaki, "Taxi demand prediction using ensemble model based on RNNs and XGBOOST," in *Proceedings of the 2018 International Conference on Embedded Systems and Intelligent Technology and International Conference on Information and Communication Technology for Embedded Systems, ICESIT-ICICTES 2018*, pp. 5–10, Khon Kaen, Thailand, May 2018.
- [5] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "DNN-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, New York, NY, USA, pp. 1–4, 2016.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016, pp. 770–778, Las Vegas, NV, USA, June 2016.

- [7] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: a graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, Dec. 2018.
- [8] H. Yao, F. Wu, J. Ke et al., "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, AAAI 2018*, pp. 2588–2595, 2018.
- [9] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 2015, pp. 802–810, 2015.
- [10] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, no. October, pp. 591–608, 2017.
- [11] G. Guo and T. Zhang, "A residual spatio-temporal architecture for travel demand forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102639, 2020.
- [12] W. Zhu, J. Wu, T. Fu, J. Wang, J. Zhang, and Q. Shangguan, "Dynamic prediction of traffic incident duration on urban expressways: a deep learning approach based on LSTM and MLP," *Journal of Intelligent and Connected Vehicles*, vol. 4, no. 2, pp. 80–91, 2021.
- [13] Y. Liu, C. Lyu, Y. Zhang, Z. Liu, W. Yu, and X. Qu, "DeepTSP: deep traffic state prediction model based on large-scale empirical data," *Communications in Transportation Research*, vol. 1, no. November, Article ID 100012, 2021.
- [14] S. Liao, L. Zhou, X. Di, B. Yuan, and J. Xiong, "Large-scale short-term urban taxi demand forecasting using deep learning," in *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*, vol. 2018, pp. 428–433, Jeju, Korea (South), January 2018.
- [15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [16] Y. LeCun and B. J. S. D. R. E. W. L. D. Boser, "Back-propagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [17] T. N. Kipf and M. Welling, "Semi-Supervised classification with graph convolutional networks," Sep. 2016, <http://arxiv.org/abs/1609.02907>.
- [18] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, pp. 1–34, 1997.
- [19] F. Chollet, "Keras." GitHub, 2015, <https://keras.io/getting-started/faq/#how-should-i-cite-keras>.
- [20] M. Abadi, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, <https://www.tensorflow.org/about/bib?hl=eng>, 2015.
- [21] D. Grattarola and C. Alippi, "Graph neural networks in TensorFlow and Keras with spektral," *CoRR*, vol. 12138, 2020, <https://arxiv.org/abs/2006.12138>.
- [22] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, San Diego, CA, USA, May, 2015.
- [23] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, p. 17, 1950, <http://www.jstor.org/stable/2332142>.
- [24] A. D. Cliff and J. K. Ord, *Spatial Processes: Models & Applications*, Taylor & Francis, Milton Park, in Oxfordshire, 1981.
- [25] L. Anselin, "Local indicators of spatial association-LISA," *Geographical Analysis*, vol. 27, no. 2, pp. 93–115, 2010.